Validating Visual Simulation of Small Unit Behavior

Dr. Amela Sadagic
Naval Postgraduate School (NPS)
Monterey, California
asadagic@nps.edu

ABSTRACT

A large number of contemporary military simulations and game-based systems employ models of human behavior where individual members of simulated military formations are represented as virtual human agents. However, we do not yet see a comparable research effort directed towards ensuring that this type of representation is realistic. While a simulation of an entire military formation has its own challenges, the realistic representations of individual humans in the same formation raises a multitude of additional issues the modelers need to be aware of. This paper presents the results of our study focused on validation of visual representations of humans and human behavior models; a specific situation examined in this work was a simulation of small unit operations in a typical urban warfare environment. Each subject in our study observed eight videos showing different actions in an urban environment, and was asked to evaluate and comment on selected performance traits in each video. Our findings suggest that two major categories of comments were raised: one dealing with the realism of human behavior (non-military component), and another dealing with the correctness of simulating military tactics, techniques and procedures (TTPs); both appear to be important when evaluating the overall realism of simulated unit behavior. Given the availability of fully immersive training systems, the increased number of trainees who get exposed to such systems, and the importance of avoiding negative training transfer, this type of system validation is becoming ever more significant. Guided by the results of this study we introduce a term 'break in behavioral presence' (BIBP) and discuss its importance in training simulations. Finally, the paper provides a basic framework for validation of human behavior models, with the ultimate goal of ensuring that the investments made in developing this type of simulation get maximized.

ABOUT THE AUTHOR

Dr. Amela Sadagic has 23 years professional research experience in computer graphics and virtual reality systems. She is currently a Research Associate Professor at the Naval Postgraduate School (NPS), Modeling Virtual Environments and Simulations Institute (MOVES), Monterey, CA, where she has been leading research on projects sponsored by the ONR, NMSO and IARPA. To date those projects have involved over 4000 USMC personnel, focusing on issues such as; smart instrumentation systems for physical training ranges, training simulations, evaluation of training effectiveness in virtual simulations, and the design of novel training methodologies and pedagogies used with virtual simulations and game based systems. In the past she was a Director of Programs at Advanced Network and Services Inc. where she designed and led programs on the use of emerging technologies in learning, and coordinated the National Tele-Immersion Initiative. Her expertise and research interests include: computer graphics and virtual environments, human factors and presence in VR, multiuser collaborative environments, game-based systems, coupling of emerging technologies with systems for training and learning, and diffusion of innovation. Dr. Sadagic holds PhD degree in Computer Science from the University College London, UK.

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE **NOV 2010**	2. REPORT TYPE	3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Validating Visual Simulation of Small Unit Behavior**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School (NPS),Monterey,CA,93943**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES **Jnterservice/lndustry Training, Simulation, and Education Cof!ference (I/ITSEC) 2010, 29 Nov ? 2 Dec, Orlando, FL.**		
14. ABSTRACT **A large number of contemporary military simulations and game-based systems employ models of human behavior where individual members of simulated military formations are represented as virtual human agents. However, we do not yet see a comparable research effort directed towards ensuring that this type of representation is realistic. While a simulation of an entire military formation has its own challenges, the realistic representations of individual humans in the same formation raises a multit11de of additional issues the modelers need to be aware of. This paper presents the results of our study focused on validation of visual representations of humans and human behavior models; a specific situation examined in this work was a simulation of small unit operations in a typical urban warfare environment. Each subject in our study observed eight videos showing different actions in an urban environment, and was asked to evaluate and comment on selected perfOrmance traits in each video, Our findings suggest that two major categories of comments were raised: one dealing with the realism of human behavior (non-military component), and another dealing with the correctness of simulating military tactics, techniques and procedures (TTPs); both appear to be important when evaluating the overall realism of simulated unit behavior. Given the availability of fully immersive training systems, the increased number of trainees who get exposed to such systems, and the importance of avoiding negative training transfer, this type of system validation is becoming ever more significant. Guided by the results of this study we introduce a tenn 'break in behavioral presence' (BIBP) and discuss its importance in training simulations. Finally, the paper provides a basic framework for validation of human behavior models, with the ultimate goal of ensuring that the investments made in developing this type of simulation get maximized.**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**11**	

Validating Visual Simulation of Small Unit Behavior

Dr. Amela Sadagic
Naval Postgraduate School (NPS)
Monterey, California
asadagic@nps.edu

INTRODUCTION

The most recent decade confirmed an undeniable and growing need for employing virtual simulations and game based systems in the military domain. They are used not only as tools in military analysis, but also as extremely powerful tools in the domain of training. For any modeling or simulation tool to be adopted and used effectively, a user community needs to have confidence that the models of the real world phenomena being simulated in these systems are as accurate as possible. Only then can they be used to augment their practice. More specifically, in the training domain, this means that virtual simulations will not stray from faithfully representing the real world so much that, when used, they would introduce a negative training transfer.

A special convenience that many virtual simulations and game based systems bring to the training domain is their ability to simulate virtual forces, sometimes called agents, autonomous agents, artificial intelligence entities or constructive elements. A small unit that needs to test its readiness to be part of a larger formation, or to act in a training setup where an opposing force is also present, most likely will not be able to recruit other humans (units) to support such training whenever they are needed. Instead, the unit will opt for training solutions that provide virtual friendly forces and virtual enemies, often together in the same training scenario. Additionally, there is a need to represent neutral populations and 'pattern of life' - passers by, local merchants, or any other characters that help present a specific cultural setup in a simulated environment that is typical for urban or village life.

The purpose of this paper is to elaborate the results of our experiment, which focused on validation of a virtual simulation representing small unit behavior, where each member of the unit is represented with a human-like virtual agent. The main objective of this experiment was to examine the facets of the validation process that were specifically tuned to systems representing autonomous virtual humans, and to provide the research community with useful and tested tools that could be used in the validation process. The specific rendition of virtual simulation used in this study was the Urban Warfare Planning Tool (UWPT), an application developed for the Behavior Analysis and Synthesis for Intelligent Training (BASE-IT) research project sponsored by the Office of Naval Research (Sadagic et al., 2009). UWPT allows users to define a 'what-if' scenario. They can craft their mission plan as a connected set of military operations and 'request' the plan be executed by simulated forces. Subsequently, they can examine how this plan gets completed by autonomous forces visualized by the UWPT. As the plan gets executed, the users can discuss and examine the extent to which the plan has, or has not been successful. The result of this process could be that the users may decide to fine-tune their original plan. They could also, for example, decide to instruct simulated forces to attack a certain building that hides an identified threat from a slightly different position or add another fire team to the attack element.

In addition to the main research objective, we set up several very specific goals for our validation study. Being that the work on functionalities and models provided by UWPT is new in the domain of virtual simulations representing individual virtual humans, we wanted to apply a validation approach that would provide us with clear pointers to areas where our current models need to be refined, and to indicate additional models that need to be developed in support of the intended functionality. Humans are extremely sensitive to representations of other humans, and while an iconic representation of an entire unit on a specific terrain represents an abstraction that is free from the lowest level of details (like the appearance and behavior of each individual unit member) and is consequently easier to model, a representation that includes a visualization of all individual members of that same unit will inevitably be exposed to the highest level of scrutiny from the human observers. Not only will the appearance and behavior of each individual agent be judged, but also the way that agent communicates and acts with other agents, and how that group reacts to the surrounding environment.

It was therefore important for us to learn what elements of small unit behavior 'stick out' and get criticized most by the human observers. The training situation in

which the level of realism becomes even more important is the one that uses fully immersive training systems. In this situation, the actions of a user may greatly depend on the extent to which the user feels as if he is in a real place (Place Illusion – PI), that the scenario played back to him is actually occurring (Plausibility Illusion - Psi), and that he is sharing that space and that scenario with other individuals (Co-presence) (Slater, 2000; Slater, 2009).

The military community recognizes the importance that these types of simulations bring to the training domain. The U.S. Marine Corps Tactics & Operations Group's (MCTOG's) Enhanced Company Operations Simulation (ECO Sim) initiative has been specifically focused on making sure that a realistic portrayal of population, insurgent, and dismounted infantry activity is present in 3D simulations used by Marines (U.S. Marine Corps, MCTOG, 2010). More detailed comment about this work will be provided in a section that follows.

This paper introduces the Department of Defense (DoD) definitions and rationale for validation of models used in simulations. A brief review of different validation methods and issues related to validation of simulations that visualize human figures is also provided. Finally, the results of our study that focused on SME evaluation of simulated small unit behavior are presented and discussed.

BACKGROUND

Definitions

Validation activity is one of several related activities prescribed by the DoD that directly concern all models, simulations, and associated data that support DoD processes, products, and decisions. The official definitions of verification, validation and accreditation are (DoDI 5000.61, 2009):

Verification: *The process of determining that a model or simulation implementation and its associated data accurately represent the developer's conceptual description and specifications.*
Validation: *The process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model.*
Accreditation: *The official certification that a model or simulation and its associated data are acceptable for use for a specific purpose.*

The same document prescribes DoD policy that includes validation (in addition to verification and accreditation): "Models, simulations, and associated data used to support DoD processes, products, and decisions shall undergo verification and validation (V&V) throughout their lifecycles." This requirement is fully justified, given the need of the military user community to have highly reliable tools and systems capable of augmenting or even replacing current work practices in that domain.

Any simulation of the complex real world processes is inevitably an approximation of the functionality and characteristics of that segment of the real world. Corresponding models are still far too coarse and unrefined to be considered as a basis for exact simulation. It is therefore more productive to see validation as "a process of increasing confidence in a model, and not one of demonstrating absolute accuracy" (Robinson, 1997). An additional issue that people working in the domains of modeling, simulation and validation need to accept is the rationale for a model that is 'accurate enough' for the intended use, i.e. it is accurate enough for the purpose for which the given model is developed and its functionality employed. 'Accurate enough' could also be defined as the model being consistent with the phenomena in the real world so that when the model is used it produces the expected results and does not introduce inaccuracies below the quality level and metrics established for the particular use. While one model may be qualified as 'accurate enough' i.e. valid for one type of use, it may not have that qualification for another type of use. Having a model that would be accurate for every possible use may not even be desired – one may need to retain a certain level of abstraction for one type of use, while a different type of use may require a very fine level of detail in the model.

Validation Approaches

Researchers and practitioners working in the domain of modeling and simulation have devised different approaches and methods to validate underlying models. Some employ objective validation using different forms of quantitative analysis; one type of objective validation is a results validation using graphical and statistical methods with well defined measures of effectiveness (Simpskin, 2001), a different objective validation approach is to use historical data (past real events) to validate simulation results (Herington et al., 2002). Other authors rely on subjective validation that involves Subject Matter Experts – SMEs, individuals who have extended knowledge of the overall domain, as well as of the intended use of the validated simulation (Goerger et al., 2005). Another type of

categorization of validation methods is white box and black box validation. White box testing requires a thorough understanding of the underlying models, while the 'black box' approach leaves all those details unknown to the executor of the validation process. Additionally, validation process can use either a bottom-up (Simpskin, 2001), or a top-down approach, or a combination of both, as in the modeling and validation of COMAND system, a theater level representation of a naval-air campaign (Herington et al., 2002).

Human appearance and human behavior that involve tactical decision-making operations are two distinct tasks that are both inherently nondeterministic. Models currently used to represent both phenomena are still to a large extent only their crude approximations. To make things more complex, military documents that describe the ways in which military operations are to be planned and executed (Tactics, Techniques and Procedures-TTP), provide only a fraction of the information the modelers need to know to simulate a military unit in a typical urban warfare setting, for example. An additional, domain-specific set of information focused on a lower level of mission planning and execution, is at times very hard or even impossible to convey in documents. The warfighters learn about them and acquire those skills as a part of a grueling regimen they go through in schools, courses they attend, and later on, during their training. Although a particular simulation may appear to respect the rules derived from the TTPs, the overall impression that the simulation leaves on human observers may still not be satisfactory i.e. "accurate enough' for particular use. In our experience, a good SME can recognize a well-organized unit just by watching the way they move in space, communicate and acknowledge each other's presence, and plan their immediate actions using non-verbal means of communication. All those cues are extremely hard to express with quantitative metrics and consequently very hard to measure using objective validation only. While objective methods for validation can and should be used for this category of simulation, the non-deterministic nature of simulated phenomena requires an additional layer of examination that has subjective validation done by the SMEs as a major component.

A large majority of the visual simulations developed for the needs of the military domain dealt with a symbolic representation of an entire unit, and its movements and actions across the space. It is only more recently that advances in developing effective virtual environments and game based systems allowed for presenting individual human figures – *avatars* - operated by real humans in real time, and *agents* - figures whose actions are completely governed by the system with no human i.e. user intervention. While the models and simulations representing the actions of the entire unit had to be validated in terms of their correct military actions as a compound unit, and the information relevant to the appearance and behavior of its constructive elements was hidden and assumed within the higher-level unit model, the simulations representing individual humans have a level of complexity an order of magnitude higher in their inner workings.

In those systems, two distinct categories of phenomena need to be modeled: one relates to the non-military characteristics, and the other one to military characteristics. Non-military characteristics assume several elements: (a) *human-like appearance*, (b) *individual behavior* - full body articulation of virtual humans including interactions with the terrain and environment; (e.g. articulated movement of head and limbs, the agent not running into walls) and (c) *team behavior* (e.g. agents not running into or through another agent). Military characteristics consist of (a) *military aspects of the warfighter's appearance* (e.g. type of uniforms and gear worn), (b) *military TTP-like behavior* (includes military doctrine, TTP, standard operation procedures – SOP), and (c) *other military behavior and phenomena* i.e. any other behavior and phenomena related to military practice that needs to be simulated for the intended use of the simulation.

The domain of Virtual Environments (VE) and Presence in VE, generated a wealth of literature focused on human perception of human-like figures i.e. avatars in VEs. Most of this literature has focused solely on basic research and abstract situations, like a small team collaboration while solving text puzzles (Slater et al., 2000); navigation and exploration in sensory-rich environments (Mehan et al., 2002), such as observing a virtual room and looking for a target letter 'written' on the walls (Pausch et al., 1997), or simply entering a virtual room and observing the situation in it (Garau et al., 2005). Fewer studies provided insights about the uses of VE technology in real life experiences from end-domains. The latter group relates to studies focused on VEs being used to study or treat phobias and other disorders, like fear of speaking in public (Pertaub et al., 2001), or post traumatic stress disorder – PTSD (Hodges et al., 2001). It is only more recently that studies focused on the effectiveness of learning and training using virtual simulations and game-based systems started to emerge. Those studies involved a fairly large number of domain (end) users like K-12 learners (Ketelhut, 2007) or military trainees as study subjects (Brown, 2010), and were concerned with real life uses and applications.

The advances in VE technologies, like the ability to render and manipulate a very large number of polygons and to allow complex user interaction in real time, as well as the development of effective approaches in modeling human behaviors, have enabled a new generation of learning and training simulations capable of representing individual avatars and agents, and complex scenarios. It is only now that we can frame the user studies with SME validation of simulations focused around real (end-domain) uses. The technology no longer represents the main obstacle to good simulation of human behaviors, and SMEs can be more effective in their validation work.

It is therefore not coincidental that users now also expect a very high level of behavioral realism and correctness when using the simulations of real world phenomena like urban warfare. One of the objectives set up by MCTOG's Enhanced Company Operations Simulation (ECO Sim) is to have a realistic model that represents a "believable level of population activity which replicates unique cultural activities" (U.S. Marine Corps, MCTOG, 2010). This particular request was related to Boston Dynamic's Dismounted Infantry Guy, DI Guy, simulation. The document clearly states the simulation objective this application needs to satisfy, as well as the need for conducting the validation effort, however it does not clarify how to go about this task. Current objective methods will be able to address tangible metrics, thus providing only one part of the necessary answer. Other methods will need to be developed to address the issues that are less tangible, with qualitative metrics that also need to be validated. Now that the technology is ready, the researchers and practitioners working on validation of simulations need to provide tested validation methodologies and a comprehensive answer about the quality of human behavior simulation. That answer is very much needed by the user community so that they feel confident in the tools they are about to use on a daily basis in their training practice.

EXPERIMENTAL DESIGN

Validation Method

The environment and simulated situations studied in our experiment were related to operations done by a small unit (fire team) in an urban warfare environment.

The approach selected for our study was to use a black box, well-structured, subjective, SME-based face validation method which utilized a visual check with pre-defined metrics. The metrics used in the study consisted predominantly of a selected set of performance traits regularly evaluated by the instructors on USMC training ranges. The decision to use the black box approach in our validation process was guided by our desire to avoid situations where participants would be too aware of the underlying models and would characterize simulated performances as 'good enough' in terms of their conformity with our selection of conceptual models rather than being 'good enough' for the particular purpose and intended use. Not knowing the details of the actual conceptual model was a better choice when we were still developing and adding new models into UWPT application. In general this approach also has a potential to produce more information on what models may still be missing from our simulation, and what elements of current incarnations of our models need to be fixed.

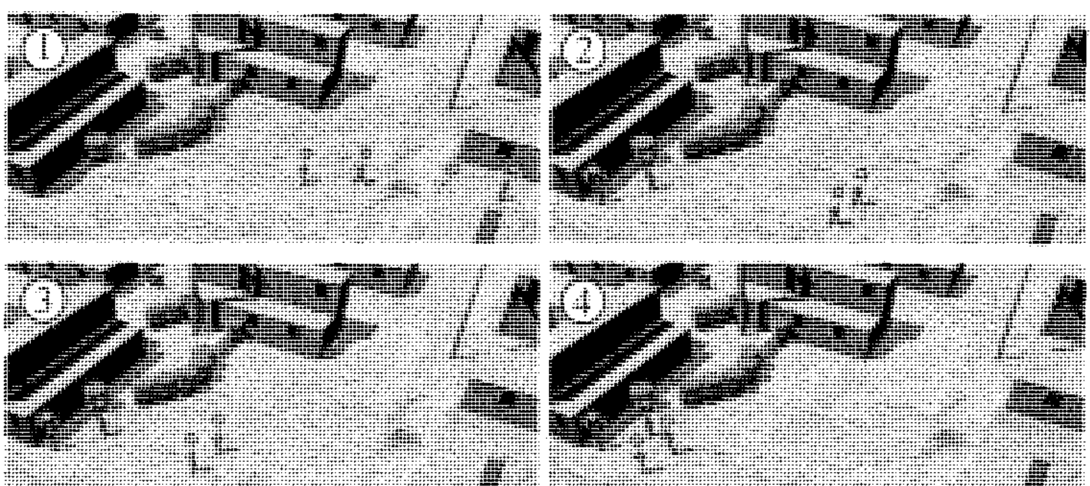

Figure 1. Bounding-movement: An example of a movement visualization evaluated in the study

As noted above, people have different views of the real world, and their understanding of the importance associated with simulated phenomena may vary as well. Our extensive observations of training exercises done on USMC ranges, and our knowledge of USMC doctrine and TTPs, suggest that there are general rules related to unit performance, and that the opinions of multiple instructors would not vary by a large extent (examples: (a) all Marines need to maintain 360 degrees security at all times, and (b) no movement or action is undertaken in a situation with a confirmed threat unless security is being provided). There are, however, situations when opinions of two instructors would differ to some extent. This is more pronounced when the instructors are asked to evaluate situations that involve tactical decision-making. With that in mind, we selected a structured validation approach with the list of performance traits regularly evaluated by the instructors on USMC training ranges, instead of opting for predominantly unstructured and open-ended validation, which is prone to higher subjective biases of SME evaluators. (A very similar rationale and approach was used in Goerger et al., 2005). Examples of performance traits that were used as metrics in our experiment include: 360 degrees security, weapon flagging - unintentionally pointing a weapon toward a fellow Marine, dispersion, hard targeting - making themselves hard targets for the enemy, movement technique when crossing a danger area, and reaction to sniper fire. A 7-point Likert scale was used for **all** metrics in this experiment.

Video Segments

Eight (8) situations were selected for evaluation; five (5) video segments were generated for each situation using our Urban Warfare Planning Tool application, making for a total of 40 video segments evaluated during this study. The 8 situations evaluated were:

1. *Scanning*: unit was stationary, scanning the environment to ensure 360 degree security,
2. *Cover-sector*: unit moved to specified position and covered a sector specified by the operator,
3. *Bounding-movement*: unit moved to new position and used bounding technique to cross danger zones (Figure 1 shows 4 stages of one such movement),
4. *Quick-movement*: unit moved quickly from its current position to a specified position,
5. *Move-and-take-position*: unit moved to a specified position in patrolling formation,
6. *Enter-the-building*: unit entered the building through the door specified by the operator,
7. *Receive-fire-and-go-firm*: unit moved to a specified position. Sniper fire was activated and unit reacted with immediate action drills.
8. *Suppressive-fire*: unit moved to a specified position and provided suppressive fire onto the sector designated by the operator.

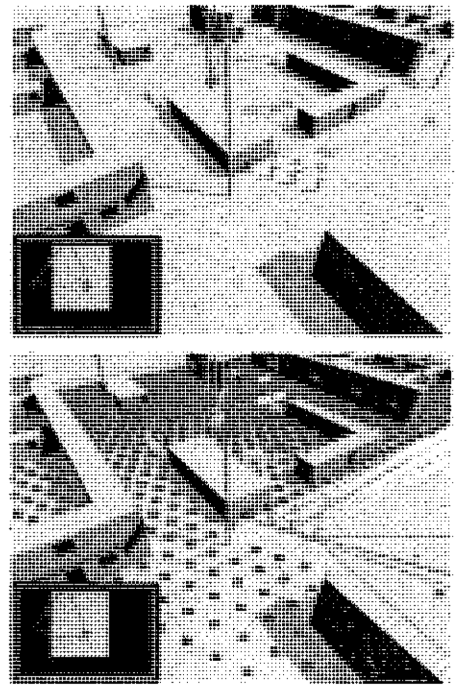

Figure 2: Cover-sector: no 'heat map' shown (first figure), and with 'heat map' shown (second figure)

All video segments were pre-generated by an operator. We wanted to exclude the impact that participants' experience with the graphical user interface could have on their subsequent evaluation of performances seen in the simulation. This also insured that participants saw exactly the same performance if they were reviewing the same video. Special care was taken to insure that all 5 video segments for one situation represented similar levels of 'difficulty' regarding the performance of the simulated unit. We also made sure they differed sufficiently so that the 5 video segments represented a solid illustration of all underlying models used to simulate a given situation. We believed that having only one video segment for one type of operation would not be sufficient to illustrate variations in simulated unit responses to each situation.

The beginning of each video showed how the operator, who was making the video, selected new positions where the fire team had to move, or how he selected a sector that the fire team had to cover. This was done to ensure that participants in the study had a good understanding of the parameters 'given' to the unit by the operator. The rest of the video showed the behavior that was generated in response to the operator's

request. In cases where it would be difficult to see the orientation of weapons in the hands of each simulated Marine, we provided several seconds of a 'heat-map' visualization as shown in Figure 2 (green color in second figure represented a segment of a terrain that was covered by multiple weapons systems). This visualization tool is part of the regular functionality in UWPT and users, if they had access to UWPT, could request it themselves.

Participants

We recruited sixteen (16) participants for the experiment, which was advertised to both faculty and students across NPS. All 16 participants were male.

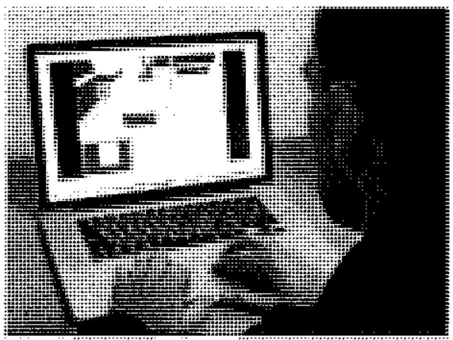

Figure 3. Participant reviews a video segment.

Procedure

At the beginning of the validation session, subjects received a standard Institutional Review Boards (IRB) documentation with a consent form that included information about the voluntary nature of their participation; the treatment of data collected during the study, including a guarantee of anonymity; as well as information about the overall experimental procedures. They were then asked to fill in a demographic questionnaire with basic information about their age, Military Occupational Specialty (MOS), years of military service, knowledge of procedures they would be evaluating, and their experience with playing video games. Participants were informed that they would be asked to review and evaluate 8 short video clips representing selected actions of a simulated small unit in an urban warfare environment, and that the 8 videos would depict 4 situations, with 2 videos for one type of situation. The decision to present 2 video segments for the same situation was dictated by our desire to have repeated exposure to the same type of performance. This would allow us to identify the frequency with which a performance was consistently evaluated as being simulated very well or simulated poorly for the same situation. The instructions clarified that participants would be able to play back each video as many times as they deemed necessary. Each participant was given a reference description for all 8 situations, 4 of which they would be viewing, and they were asked to read that information before seeing the 2 videos for that situation (videos were presented in succession). The order in which the 4 situations were presented to each participant was randomized, as was which 4 (out of 8) types of situations, and which 2 video segments (out of 5) of the same situation were presented to each participant.

The questionnaire presented after each video segment consisted of 8 questions. One question was related to 19 performance traits and 2 optional (additional) traits that participants could add if they wanted to comment on something that was not listed. Two questions inquired about the extremes in terms of participants' subjective evaluation of simulated performances: they were asked to select 5 traits of simulated performances they qualify as *Least Marine-like*, (they were also asked to say why), and another question asked them to list 5 traits they qualified as *Very Marine-like*. Four questions were related to the level of realism (overall representation of unit performance, and level of realism in individual movement), and one question was related to the team cohesion of the simulated unit.

Apparatus

All video segments were recorded in 640x480 resolution and played back on a MacBook Pro laptop using the RealPlayer application. Figure 3 illustrates our basic experimental setup and a participant who opted to enlarge his window with video play-back to make it fit the full size of the screen. Maximum screen resolution was 1920x1200.

Table 1: Basic demographic data

	Age	Years of Mil. Experience
Group GT	36	13.9
Group COM	40.6	2 civilians; mil. officers: 11.6
All	37	2 civilians; mil. officers: 12.9

RESULTS

Demographic Data

Eight (8) participants out of 16 were individuals with long military experience and expertise in ground operations (US Marine Corps and Army officers) – we call this group 'Ground Troops' – Group GT. The remaining eight (8), consisted of 'civilians and other military' – Group COM, with either military officers with MOS that was not related to ground operations

(pilots, surface naval officers) or DoD civilians. All individuals from group COM at some point in their career had multiple opportunities to become familiar with the very basic underpinnings of tactical decision-making of ground troops, and therefore, even they were not completely naïve subjects in this study. Table 1 provides information about average age and years of military experience for both groups.

Table 2: Performance traits evaluated for Bounding-Movement situations (7-pt Likert scale)

Performance	# responses	mean	st dev
Overall body movement (body shifts, body posture)	16	4.06	1.65
Keeping 360 degrees security	14	4.21	1.58
Keeping 3D security	6	2.33	1.37
Hard targeting	16	2.73	1.91
Weapon flagging	8	1.88	0.83
Gun Target Line (GTL) awareness	2	2.00	0.00
Battle-space geometry	13	2.31	1.18
Dispersion across the terrain	16	3.63	1.50
Situational awareness	9	3.67	2.12
Distribution of fires	1	1.00	0.00
Individual movement techniques when crossing danger area: bounding and bumping	11	3.64	1.75
Movement technique when crossing danger area: bounding and traveling overwatch	16	3.44	1.55
Support by fire	1	2.00	0.00
Fire and movement (maneuver) technique	2	1.00	0.00
Conducting occupied building search (entering the building)	1	1.00	0.00
Danger area crossing	14	2.93	1.33
Urban patrolling	5	2.00	1.73
Cordon & search operation	/	/	/
Reaction to sniper fire	/	/	/

Table 3: Three groups of performance traits most frequently selected as *Least Marine-like* for each situation, and a frequency with which they were selected as such

Bounding-Movement (total # comments: 59)	
Hard targeting	11
Battle-space geometry	6
Overall body movement (body shifts, body posture), Weapon flagging, Dispersion, Movement technique when crossing danger area: bounding and traveling overwatch, Danger area crossing, Urban patrolling	4
Cover-Sector (total # comments: 70)	
Hard targeting	9
Battle-space geometry, Movement technique when crossing danger area: bounding and trav. overwatch	6
Keeping 360 degrees security, Weapon flagging	5
Enter-Building (total # comments: 61)	
Hard targeting, Conducting occupied building search (entering the building)	8
Dispersion	6
Movement technique when crossing danger area: bounding and traveling overwatch	5
Move-and-Take-Position (total # comments: 65)	
Hard targeting, Dispersion	8
Individual movement techniques when crossing danger area: bounding and bumping	7
Keeping 3D security	5
QuickMovement (total # comments: 47)	
Overall body movement (body shifts, body posture), Keeping 3D security, Individual movement techniques when crossing danger area: bounding and bumping	6
Movement technique when crossing danger area: bounding and traveling overwatch	5
Hard targeting, Situational awareness, Danger area crossing	4
Receive-Fire-Go-Firm (total # comments: 74)	
Hard targeting, Reaction to sniper fire	9
Individual movement techniques when crossing danger area: bounding and bumping, Fire and movement (maneuver) technique	7
Movement technique when crossing danger area: bounding and traveling overwatch	6
Scanning (total # comments: 55)	
Keeping 360 degrees security	10
Overall body movement (body shifts, body posture), Hard targeting	8
Dispersion	7
Suppressive-Fire (total # comments: 72)	
Hard targeting	9
Individual movement techniques when crossing danger area: bounding and bumping, Movement technique when crossing danger area: bounding and traveling overwatch, Fire and movement (maneuver) technique	8
Keeping 360 degrees security, Dispersion	6

Self-reported average skill level differed between the two groups, as we expected it to. On the scale of 7 with 1 meaning 'not satisfactory', and 7 meaning 'excellent', Group GT scored all their skills consistently between 5 and 6 (mean=5.65, stdev=0.3), and Group COM scored their skills fairly low (mean=2.29, stdev=0.74) with the only exceptions for battle Space Geometry – BSG being scored as 3.3 (mean), and Situational Awareness - SA being scored as 4.7 (mean). Those two types of performances are very common for all military officers regardless of their MOS, and while participants in this group were not infantry officers themselves, they could have made critical connections and parallels between their domains (MOS) and infantry and performances in urban warfare. Most of the participants reported past experience with first-person shooter type games (13), then puzzles, strategy and card games (9), racing (8), and adventure and fantasy games (7). 11 participants

reported the use of simulations being required and used by them at some point in their military career.

Questionnaire Results

Military performance

Nineteen different performance traits were evaluated for each viewed video and the situation it presented. Table 2 illustrates how all 19 performance traits were evaluated for Bounding-Movement situations on a 7 point Likert scale (1=did not look like something that Marines would typically do at all, 7=it looked very much Marine like). The performance traits that were evaluated by the largest number of subjects, were the traits that indeed matter the most in this type of situation: 16 subjects provided their marks for *Hard targeting*, *Dispersion across the terrain*, *Movement technique when crossing danger area*, and 14 evaluated *Keeping 360 degrees security* and *Danger area crossing* (note: a subject could skip evaluating a trait if he felt it was not applicable to a given situation). Additionally, 16 subjects felt compelled to evaluate a non-military trait *Overall body movement* as well. This same trend, the subjects evaluating the performance traits most pertinent to a particular situation, has been consistent for all situations examined in our study. If we adopt the scheme where the marks 6 and 7 mean 'good', marks 4 and 5 mean 'good enough', and marks 1, 2 and 3 mean 'poor' in a domain UWPT application, we can conclude that the models dealing with *Overall body movement* and *Keeping 360 degrees security* got passing marks, and that the models contributing to other performance traits need to be perfected and some new models even added.

As an illustration of the type of specific comments participants gave about elements of Marines' movements that were *well done*, we list several comments for Bounding-Movement situations: "The cover position at the last danger area was good. The final dispersion and formation at the end of movement was excellent", "Individual movement was good in relation to independent icon action within the team", "Overall it seemed much more fluid. The general feel of the movement was not very forced."

Table 3 lists three groups of performance traits most frequently selected as *Least Marine-like* for each situation (the number listed with each group signifies the frequency with which those traits were selected as *Least Marine-like*). An important conclusion that can be derived from the results presented in Table 3 and the remaining data shown here, is the consistency with which the participants listed performance traits as *Least Marine-like* and the actual scores they gave for the same traits, where low scores were given to traits that were most listed as *Least Marine-like*. Illustrations of qualitative comments generated by participants include: (1) *Hard Targeting* (making oneself not an easy target for the enemy) in Bounding-Movement situations: "They were too much in the open, should have been up against the buildings more", "Unit stopped in areas with no cover", "Not using available nearby cover in overwatch positions", and (2) *Individual movement techniques* in Suppressive-Fire situations: "They made no use of the concrete barrier and didn't use the building for cover", "They all run across the road at the same time", "One figure will always just run back & forth for no clear reason".

Level of Realism and Team Cohesion

The level of realism is a significant parameter in any simulation of the real world. Participants in this study were asked to evaluate the level of realism for the overall representation of unit performance, and for Marine movement across the terrain, as we believed that those would matter most in the situations simulated in UWPT. Of similar significance is team cohesion. Unit operations in urban warfare are the situations where team effort is highly pronounced and mission success depends, to a great extent, on team skills and coordination. We were therefore interested to know if a simulated unit in our application gave the impression of a well organized team. In other words, were the underlying models embedded in the UWPT application good enough to simulate a well organized team, or will other models need to be added.

Table 4: Overall level of realism and team cohesion for each situation (7-pt Likert scale)

Situation	Level of realism		Team cohesion	
	mean	stdev	mean	stdev
Bounding-Movement	3.67	1.67	3.44	1.82
Cover-Sector	3.31	2.10	3.31	2.09
Enter-Building	4.38	1.63	4.39	1.69
Move-and-Take-Position	4.23	1.74	4.87	1.30
Quick-Movement	5.07	1.21	5.06	1.61
Receive-Fire-Go-Firm	3.67	1.44	3.54	1.66
Scanning	3.87	1.60	3.75	1.57
Suppressive-Fire	3.38	1.39	3.25	1.44

Table 4 illustrates the results for the overall realism and team cohesion - the extent to which simulated units were qualified as well organized and coordinated teams in each situation (7 point Likert scale: 1=representation did not look realistic at all / they did not look like a well organized and coordinated team, and 7=representation looked very realistic / they looked like a very well organized and coordinated team). It is

interesting to note that situations involving less complex general movement of simulated units across the terrain (Quick-Movement, Move-and-Take-Position, Enter-Building) scored higher for both characteristics than situations where more complex actions were expected. This is well aligned with other results in our study - the subjects were more rigorous in evaluating simulated situations where the threat from the enemy was more immediate and the level of threat higher, as well as situations that required more complex unit responses with multiple actions being done simultaneously. It also suggests that our models of general movement across the terrain were 'good enough' for situations with lower threat level, but were not 'good enough' in simulating multiple actions taking place simultaneously in situations with higher threat levels. Quick-Movement situations, for example, were qualified as the situations with very high levels of realism and high team cohesion (Table 4), and their performance traits were selected fewest times as 'least Marine-like' (Table 3).

Free Observation

Participants were not requested to report how many times they played-back each video. However, we did ask about this at the end of the session, and they reported reviewing some videos 3-4 times (especially at the beginning of the session), and some only once. It has been noticed that participants did not rush through the video segments but instead took time to review each video thoroughly and only then provided feedback. Total time to review 8 videos and fill out the questionnaires ranged between 60 and 80 min. We also observed that about half of the participants opted to maximize the viewing window so that the video play-back filled the entire screen (Figure 4).

Discussion

The results of our analysis suggest several areas in need of improvement:

- Movement model: both movement of individual agents and movement of the entire unit. Movements in situations with lower threat level were 'good enough', however, movements and actions in situations with higher threat level were not satisfactory.
- Interaction of unit with their immediate environment (hard targeting, use of cover): This was highly scrutinized and scored the lowest marks. New models that extensively use micro-terrain features need to be developed and integrated.
- Tighter connection with TTPs. Example: entering a building is currently done in a very rudimentary way, with no proper stacking formation.
- Team cohesion and model of team collaboration. We propose introducing a more complex model of team cognition as well as elements of non-verbal team communication (hand signals and gestures).

The study results we have obtained so far have allowed us to test our approaches and identify areas where improvements are very much needed. The validation work in general requires more complex statistical analysis than what we were able to conduct here - in this study some individual videos were seen by only 2 participants, while others were seen by 5. In order to draw more general conclusions a much larger data set needs to be obtained - this is our goal for a second round of validation sessions with new SMEs.

Breaks in Behavioral Presence (BIBP)

After reviewing the comments made by participants about the performances that were *Least Marine-like*, we have a very good basis to conclude that certain elements of simulated performances were extremely powerful in terms of 'sticking out' and providing constant reminders that units represented in the videos were nothing more than computer programs. Similar to the term 'break in presence' (BIP) that is used in VE literature to characterize phenomena from the real world that interfere with a simulated illusion of a virtual world (Slater, Steed, 2000), we identify the '*breaks in behavioral presence*' (BIBP) as a set of artifacts in simulations of human behavior – the imperfections in a simulation that are powerful enough to diminish the overall impression of the simulated behavior and in extreme cases, disrupt the basic task that the simulation is trying to achieve.

Basic Framework for Validation of Human Behavior

Our results and our findings suggest that as part of a basic framework for validation of simulations that include behavior of individual human figures, a formal subjective validation done with SMEs needs to be included along with an objective validation. We believe this should be done at a minimum of two points in the process of developing the simulation. First it should be done when all models are put together by the developers. This validation will provide important pointers related to the imperfections in the current models and indicate what other models may need to be added. A second validation should be performed at the very end when the simulation needs to be officially validated and receive its final seal of approval.

CONCLUSIONS

Validation of simulations that visualize individual human figures acting in desired situations and with desired level of realism, will be a prominent research topic in the coming years. The research community will have to adopt a validation framework that is capable of addressing this topic both effectively and reliably. Given the results of past research in the VR/VE modeling and simulation community, we believe a comprehensive set of validation methods will need to be available to complete that task. One of those methods will almost certainly be a validation that relies on the knowledge and expert opinions of SMEs.

ACKNOWLEDGEMENTS

The authors would like to acknowledge our sponsor, The Office of Naval Research, and transition customers: The Program Manager for Training Systems (PM TRASYS), and Training and Education Command (TECOM), for their unreserved support in our work. Special thanks go to the students and colleagues from NPS who volunteered their time and participated as subjects in this study. The opinions expressed here are those of the author and do not necessarily reflect the views of the sponsor or the Department of Defense.

REFERENCES

Brown, B. (2010). A Training Transfer Study of Simulation Games. *Master thesis, Naval Postgraduate School, Mar 2010.*

DoDI 5000.61 (2009). DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)

Garau, M., Slater, M., Pertaub, D.-P., Razzaque, S. (2005). The Responses of People to Virtual Humans in an Immersive Virtual Environment, *Presence: Teleoperators & Virtual Environments, February 2005, vol. 14, no. 1,* pp. 104-116

Goerger, S.R., McGinnis, M.L., and Darken, R.P. (2005). A Validation Methodology for Human Behavior Representation Models. *The Journal of Defense Modeling and Simulation: Applications, Methodology, technology 2005, 2(39).* pp. 39-51

Hodges, L., Anderson P., Burdea, G., Hoffman, H., & Rothhaum, B. (2001). Treating Psychological and Physical Disorders with VR. *IEEE Computer Graphics and Applications,* pp. 25-33.

Herington, J., Lane, A., Corrigan, N., & Golightly, J. A. (2002). Representation of historical events in a military campaign simulation model. *In E. Yücesan, C.-H. Chen, J. L. Snowdon, & J. M. Charnes (Eds.), The 2002 Winter Simulation Conference.* pp. 859-863.

Ketelhut, D.J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *The Journal of Science Education and Technology, 16 (1), 99–111.*

Meehan, M., Insko, B., Whitton, M., Brooks, F.P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics (TOG) , Proceedings of the 29th annual conference on Computer graphics and interactive techniques, Volume 21 Issue 3*

Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying Immersion in Virtual Reality. *Proceedings of SIGGRAPH, 1997,* pp. 13-18.

Pertaub, D-P., Slater, M., Barker, C. (2001). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence: Teleoperators and Virtual Environments, 2001, 11(1),* 68-78.

Robinson, S. (1997). Simulation Model Verification and Validation: Increasing The Users' Confidence. *Proceedings of the 1997 Winter Simulation Conference, ed. S. Andradóttir, K. J. Healy, D.H. Withers, and B.L. Nelson.* pp. 53-59

Sadagic, A., Welch, G., Basu, C., Darken, C., Kumar, R., Fuchs, H., Cheng, H., Frahm, J.M., Kolsch, M., Rowe, N., Towles, H., Wachs, J., and Lastra, A. (2009). New Generation of Instrumented Ranges: Enabling Automated Performance Analysis. *Proceedings of 2009 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC-2009), Orlando, FL.*

Simpkins, S.D., Paulo, E.P., & Whitaker, L.R. (2001). Case study in modeling and simulation validation methodology. *The 2001 Winter Simulation Conference,* Piscataway, NJ: Institute of Electrical and Electronic Engineers, pp. 758-766.

Slater, M., Sadagic, A., Usoh, M., and Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study, *Presence-Teleoperators and Virtual Environments, 9(1),* pp. 37-51.

Slater, M. & Steed, A. (2000). A Virtual Presence Counter. Presence: *Journal of Teleoperators and Virtual Environments. 9 (5),* 413-434.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society, Biological Sciences, Dec 2009, 364 (1535).* pp. 3549-3557

U.S. Marine Corps, MCTOG (2010). Enhanced Company Operations Simulation (Eco Sim) Initiative, 2 Feb 2010.